

ERIC Document Resume		See Instruction Sheet		DATE OF RESUME 6/18/65	DATE RECEIVED
A. PUBLICATION DATA					
1. ERIC NUMBER ED000077		2. ORIGINATOR CONTROL NO(S) 7415		3. IDENTIFICATION NO. N.A.	
4. AUTHOR(S): Cobb, William E.				DO NOT WRITE IN SPACE BELOW	
5. TITLE OF DOCUMENT ABSTRACTED: An Evaluation of Graduate Students' Knowledge in Professional Fields of Education as Revealed by a Diagnostic Examination				MICROFILM COPY AVAILABLE YES ☒ NO ☐	
6. PUBLICATION TITLE: Penn State Review of Educational Research				COPYRIGHT GRANTED YES ☐ NO ☒	
7. EDITOR(S): N.A.				COPYRIGHT RELEASE FILED YES ☐ NO ☐	
8. PUBLISHER OR SOURCE The Pennsylvania State Univ.				TYPE OF REPORT PROPOSAL ☐ IN-PROGRESS ☐ FINAL ☒ FOLLOW-UP ☐	
9. PLACE OF ORIGIN The Penn State Univ. ST. University Park, CITY, STATE Pa.				OTHER ERIC REFERENCES:	
10. VOL. V	11. PAGES 65-67	12. MO./YF. Nov. 53	13. TOTAL PP. 3	TELE. INST/LCC CODE	EXT. ERIC SATELLITE CODE
B. INDEXING DATA					
14. SUBJECT AREA			15. TYPE OF ENDEAVOR abstract of doctoral dissertation and experimental study		
16. INDEXING TERMS (8-12)					
experimental studies		measurement		students	
abstracts		test construction		professional edu.	
examinations		graduate edu.		teacher edu.	
17. ABSTRACT (MAX. 250 WORDS)					
An objective analysis of the confidence that can be placed in the validity and reliability of the Diagnostic Examination. The study further analyzes the value of the individual test items so that weak areas may be improved.					
ERIC No.					
18. SOURCE OF ABSTRACT					
C. FUNDING DATA					
19. SUPPORTING ORGANIZATION			20. DIVISION OR DEPT.	21. CONTRACT NO.	
			22. TOTAL COST OF PROJECT	PROJECT NO.	
			23. DURATION OR PROJECT	U.S. Department of Health, Education, and Welfare Office of Education	

5/Reported

Penn State Review of Educational Research
Nov 1953
Vol II No 1

7415

AN EVALUATION OF GRADUATE STUDENTS' KNOWLEDGE
IN PROFESSIONAL FIELDS OF EDUCATION
AS REVEALED BY A DIAGNOSTIC EXAMINATION

- William E. Cobb -

Summary:

HE 7415

1. Origin of and Need for Study. The Diagnostic Examination has been used in the Department of Education of the School of Education of The Pennsylvania State College since 1950. In order to obtain the maximum efficiency from any instrument it is necessary to know what amount of confidence can be placed in the instrument. The amount of confidence that can be placed in the Diagnostic Examination has been the object of much subjective conjecture by both faculty and students. This study is an objective analysis of the confidence that can be placed in the validity and reliability of the examination. It further analyzes the value of the individual test items so that weak areas may be remedied.

2. Review of Literature. The literature on this subject disclosed two important facts. In the first place there is an almost universal agreement among the authorities in the field of testing that diagnostic testing is a desirable aspect in rendering advisory service to students. In the second place no commercial standardized tests or scales, which were designed to do the tasks assigned to the Diagnostic Examination, were located.

3. Procedures. The scores obtained from a group of 579 candidates were selected. Several statistical devices were used in obtaining objective evidence concerning the Diagnostic Examination. To determine the reliability of the categories and of the total test, the Kuder-Richardson formulas were used. To determine the validity of the categories and of the total test, three methods were used. First, the curricular validity was investigated. Second, the critical ratio techniques were used to determine whether significant differences existed between the averages of those testees selected by various advisers as superior and the averages of those testees not so selected. Third, the critical ratio techniques were used to determine whether significant differences existed between the averages of the 579 masters candidates and the averages of 100 doctoral candidates. To make an item analysis, two devices were used. First, the percentage passing each item was obtained. Second, the point biserial coefficient of correlation between each item and the score of its category was obtained. These two were combined into an index and a rating of each item was made.

4. Reliability. The Diagnostic Examination is weakest in the area of reliability. Only on the total test and on the Art Education Category is a reliability of 0.85 or better obtained. However, if the Thurstone criterion for reliabilities to be used with profiles are adopted, then the user of the test scores is furnished with adequate rules for putting confidence in the categories. In the categories of Administration and Supervision, Methods, Philosophy, Advanced History and Philosophy, and Guidance the conclusions given by the test results will be error once in

U. S. DEPARTMENT OF HEALTH, EDUCATION AND WELFARE

Office of Education

This document has been reproduced exactly as received from the person or organization originating it. Points of view or opinions stated do not necessarily represent official Office of Education position or policy.

ED 0000 -65- 077

eight times. In the categories of Art Education, Educational Psychology and in the Total Test the results will be in error not more than once in twenty times. This is the five per cent level of confidence. These ratios should be considered in using the results of this examination.

5. Validity. All three of the criteria used to establish validity support the validity of the Diagnostic Examination. Face or Curricular Validity was established by expert opinion. In both measures of statistical Validity the categories of Administration and Supervision, Philosophy, Audio-Visual Aids, Statistics and Research, Business Education, Guidance and Educational Psychology showed significant differences at the one per cent level of confidence. Two other categories, Special Education and Art Education, showed significant differences at the one per cent level on one measure and at the five per cent level on the other. Three other categories, Methods, Advanced History and Philosophy and Higher Education, showed significant difference at the one per cent level on the other measure. In the case of two categories, Higher Education and Advanced History and Philosophy, the failure to show a difference on one measure may be attributed to the relatively small number of candidates for the master's degree who selected these courses. From these facts the validity of the categories and of the total test is inferred.

6. Item Analysis. All items with two exceptions, item number 24 in the Art Category and item number 81 in the Guidance Category, showed positive point biserial correlations with the score for their particular category. This signifies that each of the other two hundred ninety-eight items made contributions to the total score of their category. The amount of these contributions differed considerably as seen by the cumulative index. Since they had the highest cumulative index, the most valuable items were item number 68 and item number 72 in the Audio-Visual Aids Category, item number 39 in the Art Category and item number 110 in the Educational Psychology Category. It is also to be noted that certain items made insignificant contributions to the scores on the particular category in which they were placed. The number of these items found in their respective categories are Administration and Supervision 1, Methods 2, Philosophy 2, Audio-Visual Aids 2, Statistics and Research 2, Special Education 2, Business Education 2, Advanced History and Philosophy 2, Art Education 11, Guidance 8, Educational Psychology 7, and Higher Education 2.

7. Use of Scoring Formula. In all categories except Art Education the use of a scoring formula would not change the scores of more than four per cent of the testees by more than one score point. In fact in seven categories less than one per cent of the testees would be affected to this degree. However, in the Art Category the scoring formula would change the scores of 23.8 Per cent of the testees by more than one raw score point. It is further seen that the scoring formula, $S = R + \frac{0}{n}$ would be the most useful if ease in scoring is taken as a criterion.

Conclusions.

First Conclusion: No other test or examination gives the information that is obtained from the Diagnostic Examination.

Second Conclusion: The Diagnostic Examination can be improved by a

revision. Such a revision could improve the reliability of the test to the point where each category could be used reliably at the five per cent level of confidence. In order to do this it is necessary to assume that new items of a caliber similar to the ones now in use could be written. The changes necessary to accomplish the desired goals would include the elimination of poor items and a change in the total number of items allotted to each category. The items to be eliminated and the desirable length of each category are listed in the study.

In this assignment any change in the construction of an item is considered to be the same as the deletion of the original item and the substitution of a new item. The above system of revising the test retains 300 as the total number of items. The reliability of each category would be increased approximately to a 0.60. Then by changing the profile to include cut-off points one sigma above and below the mean value the reliability of each category would be increased to the five per cent level of confidence. If any categories were not to reach the desired reliability coefficient of .60, then the replacement of items with cumulative indices of 1 with items having higher indices would produce the same result.

Third Conclusion: To raise the reliability of each category to a minimum reliability coefficient of 0.85 or better would necessitate a change in the form of the test if the present time limit is to be maintained. Two possibilities are suggested: Vocabulary type items might be used inasmuch as this type of item requires less time to answer and more items could be included. The other possibility would be the use of impure items. Impure items would contribute weighted values to each of several categories.

Recommendations:

First Recommendation. The present Diagnostic Examination should be used until a better substitute is found. Such a substitute could be a new test or a revision of the present one. If the current edition of the Diagnostic Examination is used, then a scoring formula of the type, $S = R + \frac{O}{N}$, should be used for the Art Education Category. All other categories would be scored by the number correct. The total score would be a sum of the scores of the individual categories.

Second Recommendation. The present issue of the Diagnostic Examination should be revised as soon as the opportunity presents itself. The revision should embody the conclusions reached in this study. A valuable addition to this revision would be the inclusion of thirty additional items. Ten of these items should be additional items for one of the categories of the test. A second ten could be related to another category. Those items of this group which showed highly cumulative indices would be placed in the third set of ten items in the next administration of the test. In this way a constant source of new items for improvement of the test would be available. It should be noted that the thirty additional items would not be considered in the student's score. A judicious selection of the categories for which items should be added would insure the constant improvement of the weaker categories.